RSAC 2025 Conference

Many Voices.
One Community.

SESSION ID: PART4-W09

Only Regulators in the Building - EU AI Act, Compliance, Trust and AI

MODERATOR: **Prof. Dr. Norbert Pohlmann**
Director Institut für Internet-Sicherheit - if(is), Gelsenkirchen, Germany
President of the board, TeleTrusT; Berlin, Germany

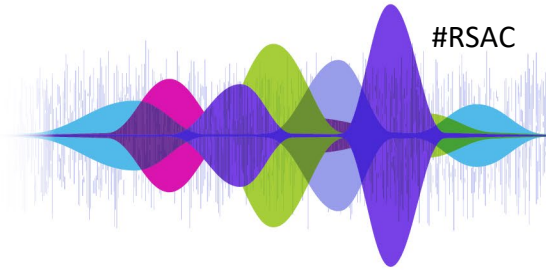
PANELISTS:

Ulla Coester
Project Director Trustworthiness
Institut für Internet-Sicherheit - if(is), Gelsenkirchen, Germany

Dr. Kim Nguyen
SVP Innovations
Bundesdruckerei GmbH, Berlin, Germany

Dr. Christian Schläger
CEO
Build38, Munich, Germany

Topics

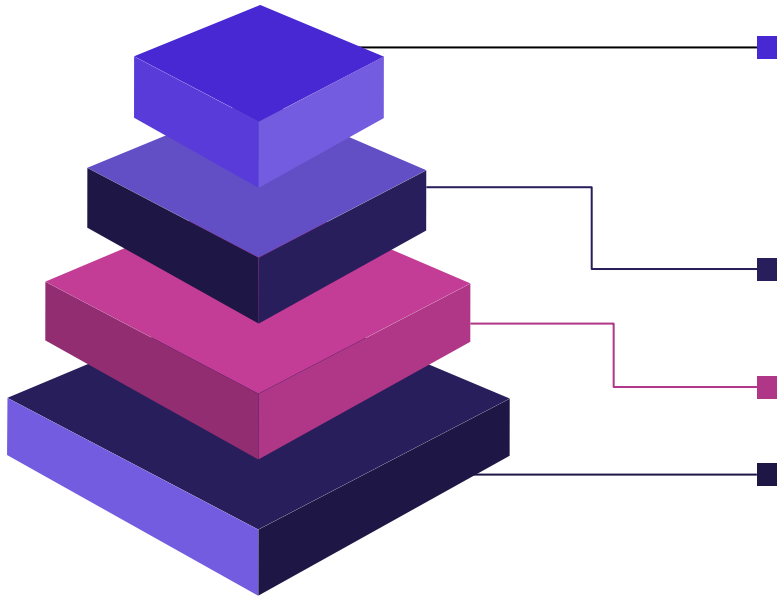


- AI regulation: its content and timeline
- AI and trustworthiness: how to define and establish trust in AI
- AI and data: how trust in data will play a vital role in AI
- AI and cybersecurity

Season 1: The EU AI story…

EU Artificial Intelligence Act (AI Act)

- What is the EU AI Act?
  - It regulates the use of AI systems according to their risks and creates rules for developers and companies that use or offer AI in the EU
  - Aim: Ensure that AI is used safely, transparently and ethically while providing a framework for innovation
- Risk categories:



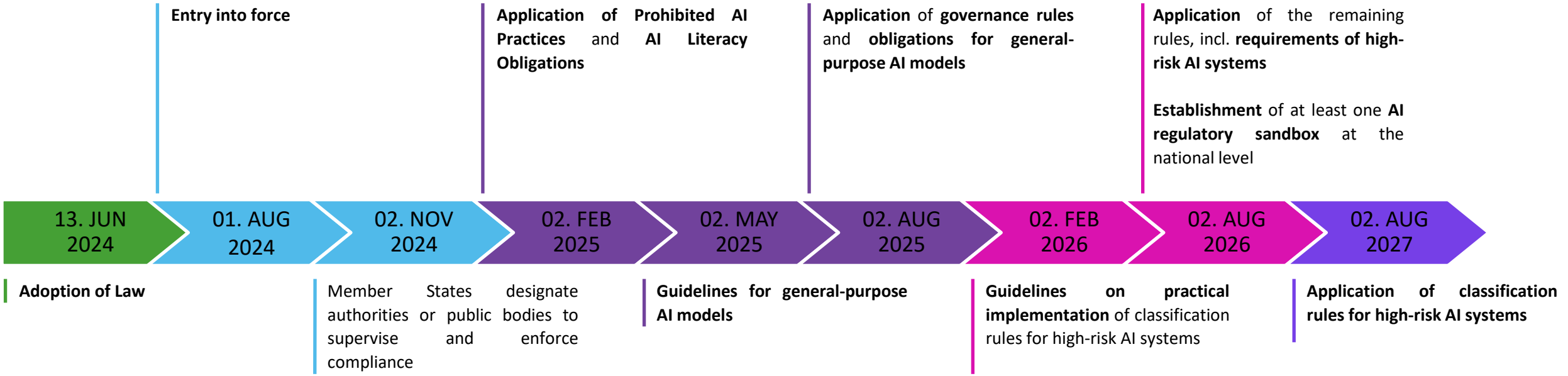
Unacceptable Risks: All AI systems considered a clear threat to the safety, livelihoods and rights of people are banned.

High Risk: Subject to strict obligations before they can be put on the market.

Limited Risks: Permitted with transparency requirements.

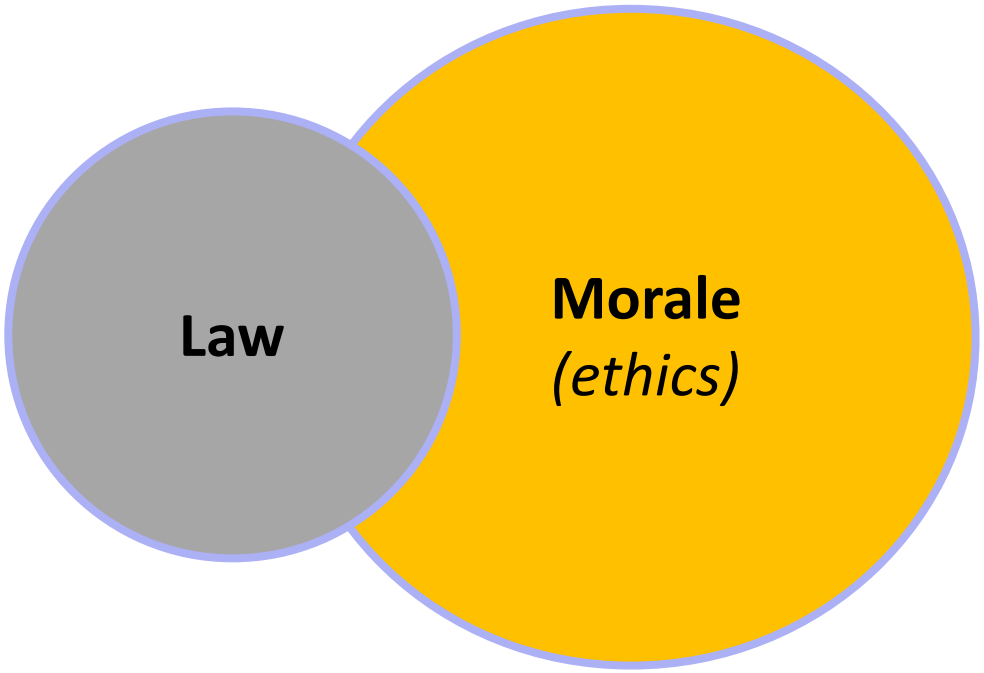
Minimal Risks: Permitted in compliance with applicable law.

Timeline - Application

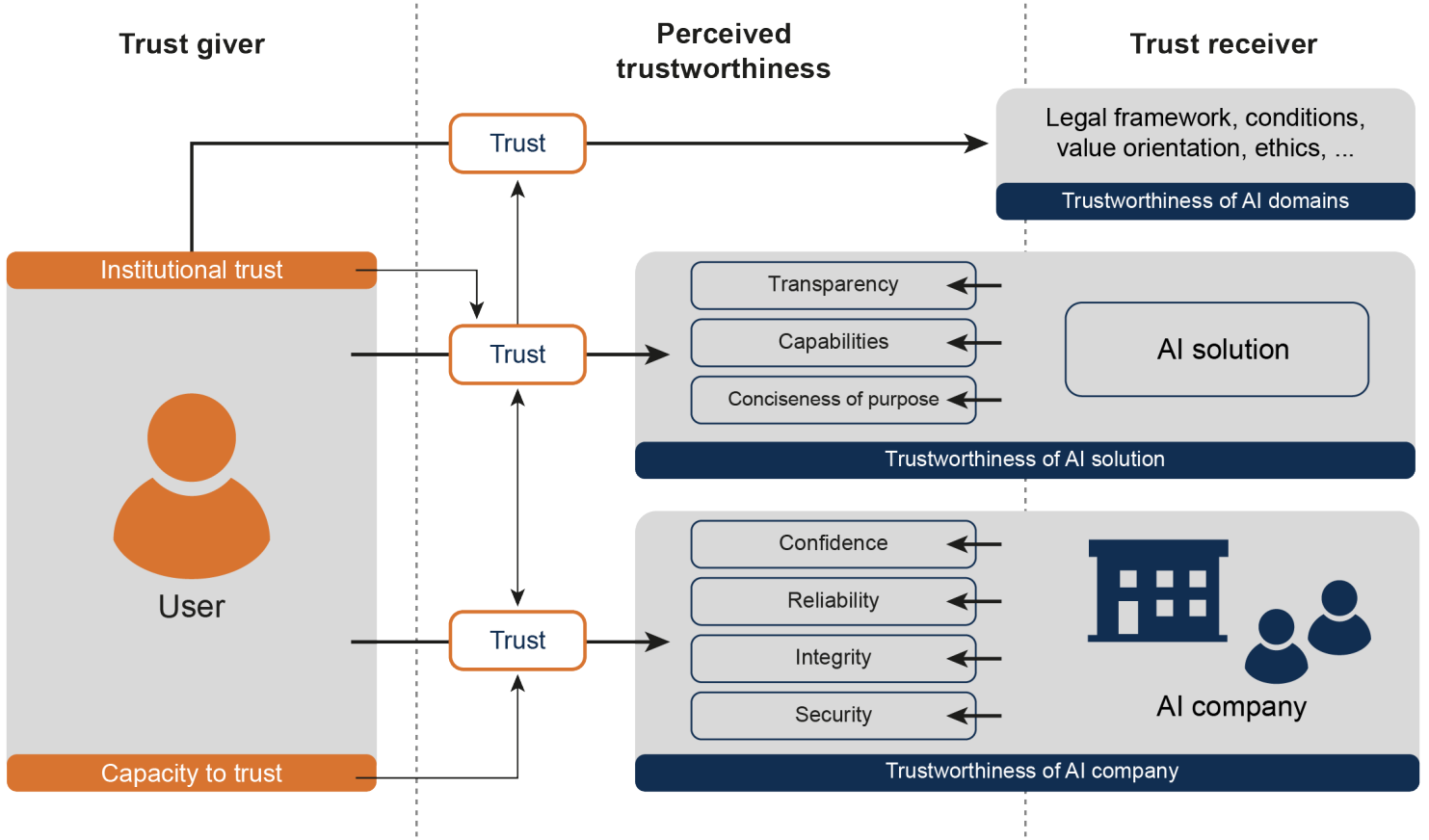

13. JUN 2024
Adoption of Law
01. AUG 2024
Entry into force
02. NOV 2024
Member States designate authorities or public bodies to supervise and enforce compliance
02. FEB 2025
Application of Prohibited AI Practices and AI Literacy Obligations
02. MAY 2025
Guidelines for general-purpose AI models
02. AUG 2025
Application of governance rules and obligations for general-purpose AI models
02. FEB 2026
Guidelines on practical implementation of classification rules for high-risk AI systems
02. AUG 2026
Application of the remaining rules, incl. requirements of high-risk AI systems
Establishment of at least one AI regulatory sandbox at the national level
02. AUG 2027
Application of classification rules for high-risk AI systems

Trustworthiness

Code of Conduct



The Foundation of Trustworthiness



Model of Trustworthiness

In the context of digitalization, trustworthiness is an essential attribute – it is a significant differentiating factor.

A Trusted *Eco-System*

Trustworthiness-Platform

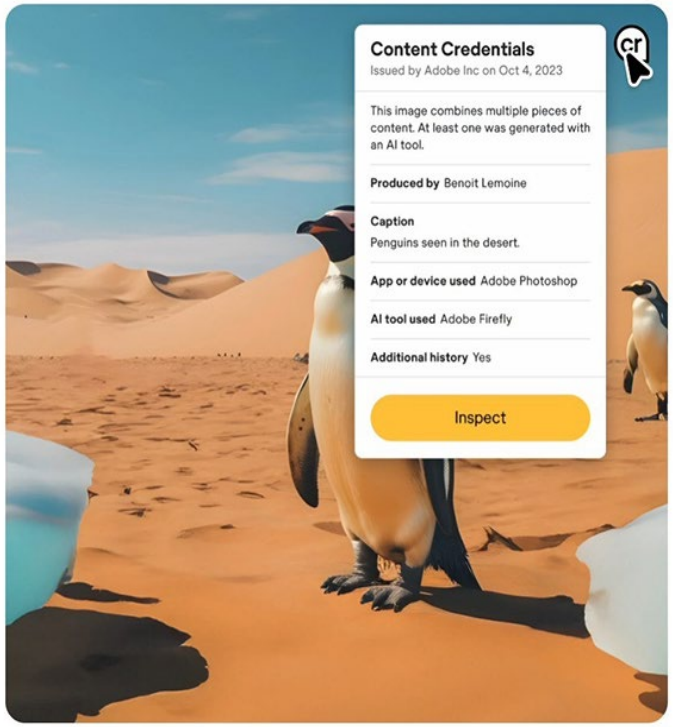
Domain-Knowledge + AI-Provider and their Solutions + User Companies

Content Credentials ensure content authenticity in times of increasing AI generated content

Source: wikipedia.org

Deepfakes and synthetic media are increasingly difficult to distinguish from real content

This is a **growing threat for content authenticity** and the trust of the public in media

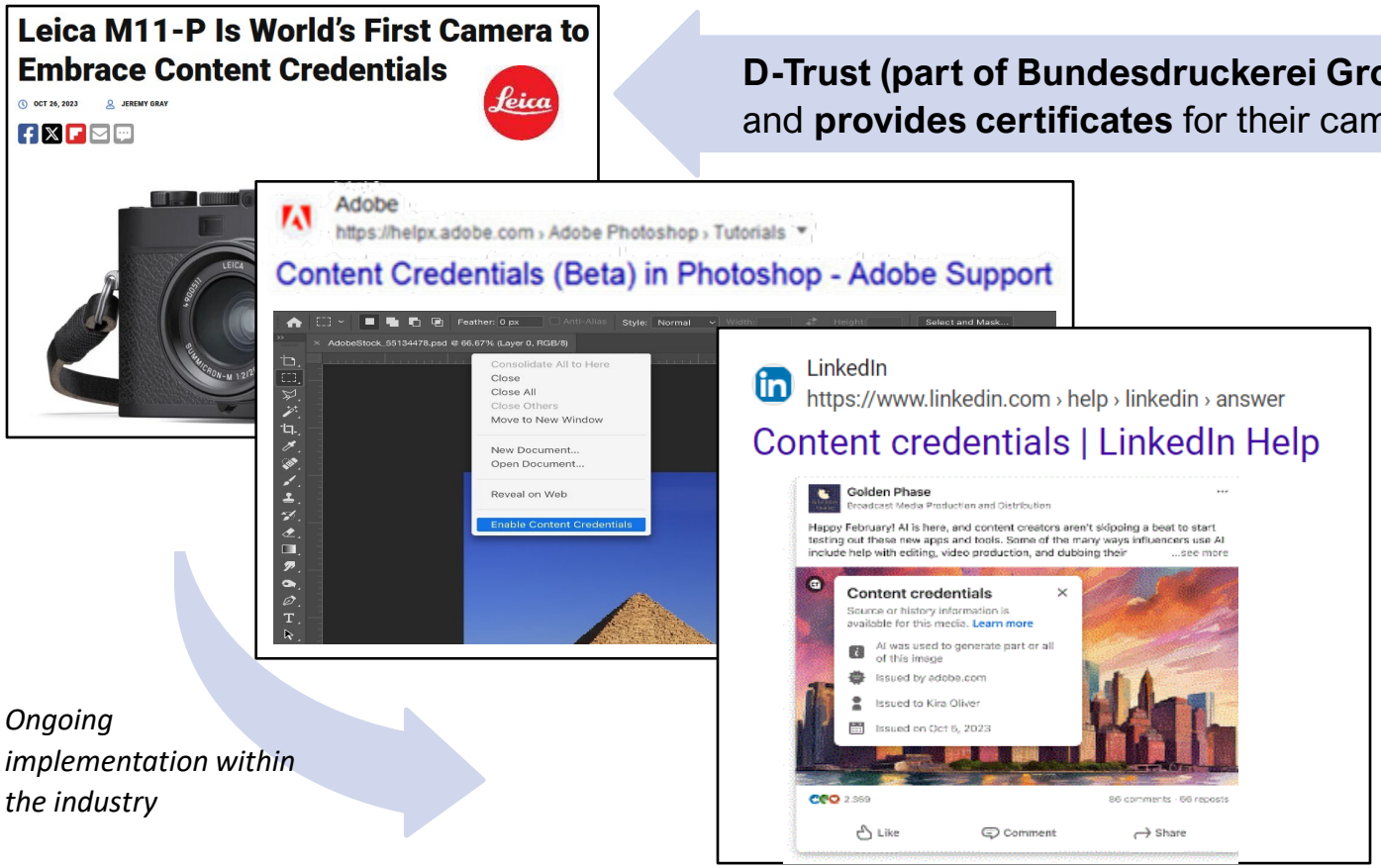

Content Credentials
Issued by Adobe Inc on Oct 4, 2023
This image combines multiple pieces of content. At least one was generated with an AI tool.
Produced by Benoit Lemoine
Caption
Penguins seen in the desert.
App or device used Adobe Photoshop
AI tool used Adobe Firefly
Additional history Yes
Inspect

Source: adobe.com

Meta data of the original content and **editing history** are stored in Content Credentials

Certificate based signing of the Content Credentials ensures tracebility of content creation and editing

Adaption of Content Credentials within the content creation industry is growing



Ongoing implementation within the industry

D-Trust (part of Bundesdruckerei Group) operates a CA for Leica, and **provides certificates** for their cameras with Content Credentials

Standardized by the **„Coalition for Content Provenance and Authenticity"** (C2PA) and promoted by the **„Content Authenticity Initiative"** (CAI) with over 4000 members

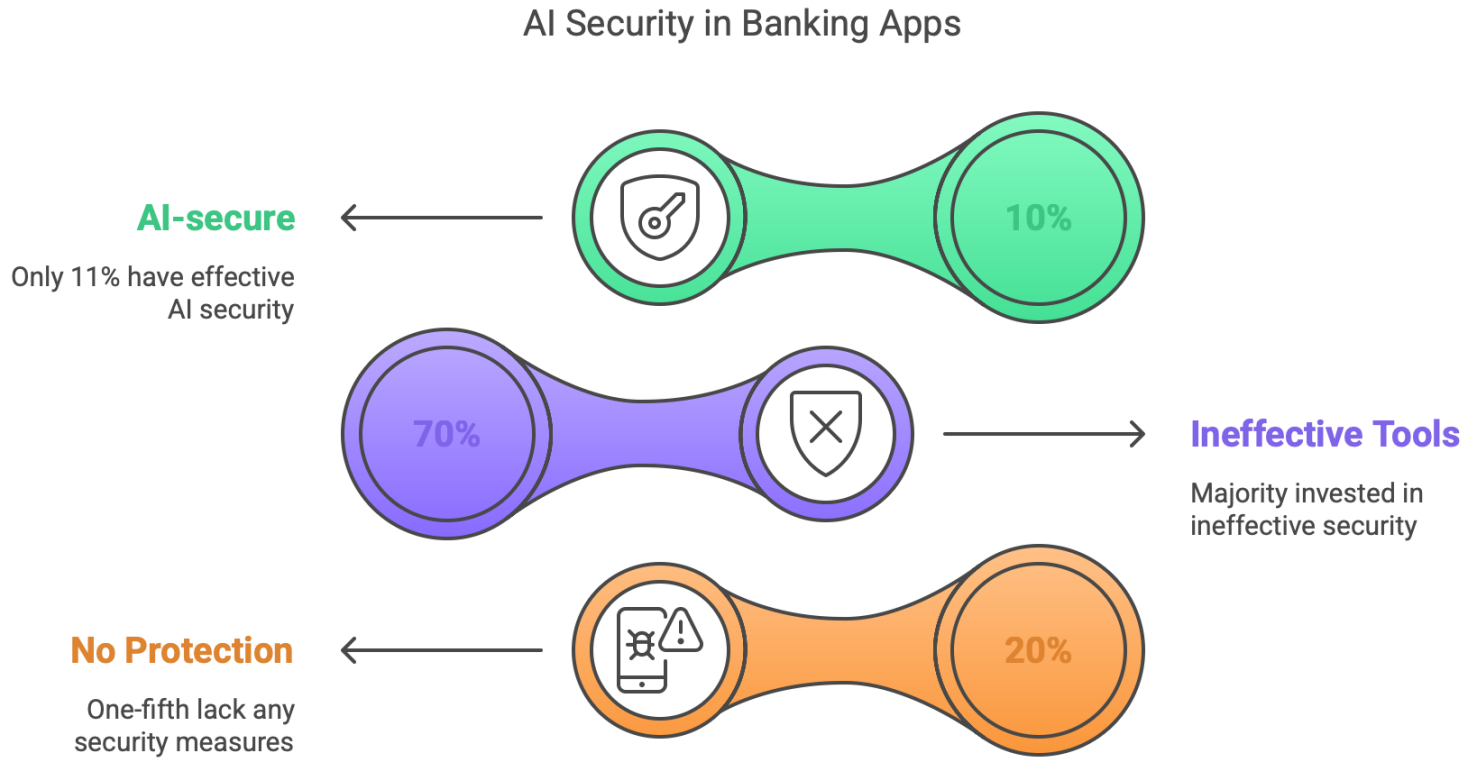
C2PA Steering Committee Members

Adobe, amazon, BBC, Google, intel, Meta, Publicis Groupe, SONY, truepic, Microsoft, OpenAI

« Soft-binding » can be implemented using **International Standard Content Code** (ISCC) acc. to **ISO 24138:2024**.

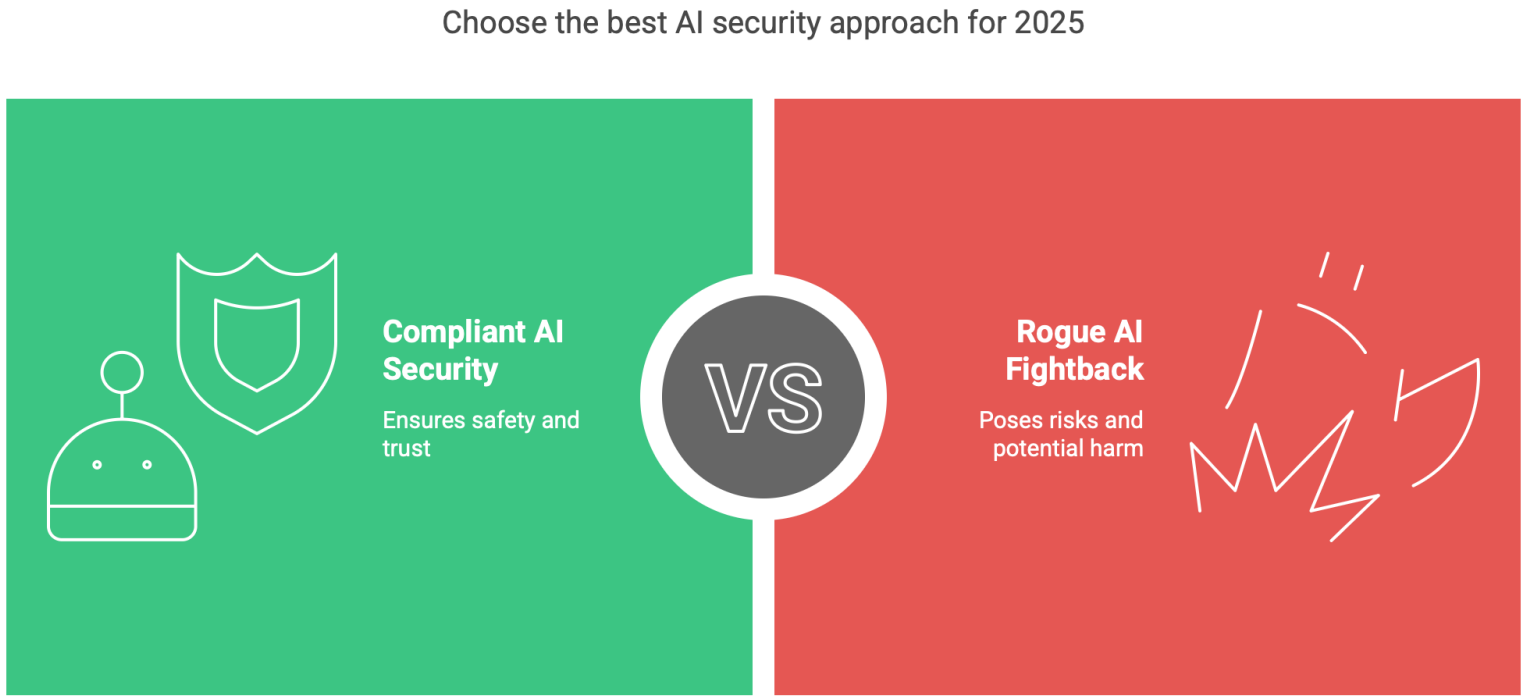
AI in the Business World Today
The AI Security Gap

- AI is mainly a source of threats – not of solutions in cyber security
- A 2024 study of 200 banking Apps in Europe reveals that only 10% have ample protection
- AI is an attack enabler
- 0,2% of all banking transaction requests to the API are with malintent – AI is pushing this number up – and the intensity of the attack



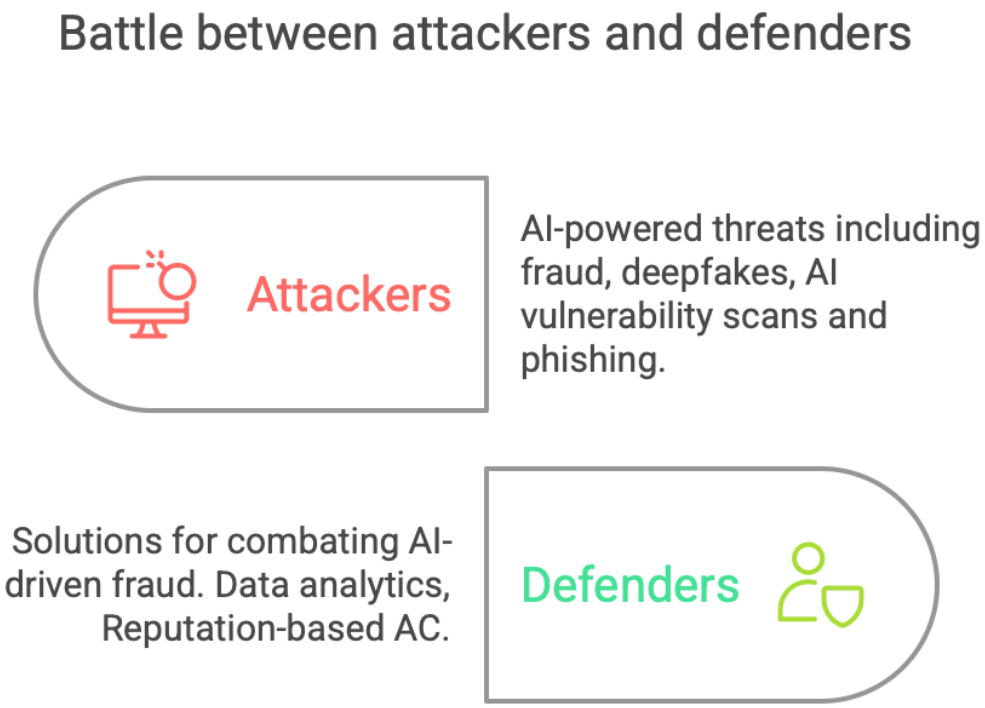
The EU AI Act and Industry Impact
The Dilemma of AI Security

- Companies choose how to respond to the AI threats – as respond they must:
- Leveraging AI within reason to strengthen cybersecurity or go rogue
- Every company aims at
  - Minimizing false positives
  - Avoid AI-driven discrimination
  - Protect customer privacy
  - Have transparency
- Providers must adhere to these wishes


Choose the best AI security approach for 2025
Compliant AI Security
Ensures safety and trust
VS
Rogue AI Fightback
Poses risks and potential harm

Who Will Win?
Trustworthy AI vs. AI-Driven Attacks

- Trustworthy AI Matters to Customers
- European businesses increasingly demand security solutions that emphasize:
  - High-quality datasets free from bias and inaccuracies.
  - Human oversight mechanisms to mitigate AI risks.
  - Consistent AI performance and reliability in threat detection.
  - Data protection compliance in line with GDPR
- Cyber Security Providers adhering to these requirements (as included in the EU AI Act) have a competitive advantage


Battle between attackers and defenders
Attackers
AI-powered threats including fraud, deepfakes, AI vulnerability scans and phishing.
Solutions for combating AI-driven fraud. Data analytics, Reputation-based AC.
Defenders

#RSAC

Visit the German Pavilion at North Expo Hall, Booth 5469

The Partners at the German Pavilion at RSAC Conference 2025

AUCONET Business Infrastructure Control

betasystems

m2trust

KOBIL

MACONIA

enclaive

itWatch GmbH

LocateRisk

MALTEGO

MTRIX® GmbH [AUTHENTICATION SOLUTIONS]

ONEKEY

TÜVIT TÜVNORDGROUP

TeleTrusT Pioneers in IT security. Bundesverband IT-Sicherheit e.V. IT Security Association Germany

RSAC™ 2025 Conference

Many Voices.
One Community.

Thank you for your attention!
Questions?